

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2005, 10:51:32 ; Search time 163 Seconds  
(without alignments)  
28.473 Million cell updates/sec

Title: US-10-618-336A-6  
Perfect score: 70  
Sequence: 1 MCPTDMPASLCM 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: geneseqp1980s:\*

2: geneseqp1990s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

%

Result	Query					Description
No.	Score	Match	Length	DB	ID	
1	44	62.9	121	4	AAB95737	Aab95737 Human pro.
2	44	62.9	296	8	ADJ50939	Adj50939 Human nov
3	44	62.9	307	8	ADJ50941	Adj50941 Human nov
4	44	62.9	329	8	ADM87220	Adm87220 Human pro
5	44	62.9	351	5	AAE18955	Aae18955 Human cel
6	44	62.9	351	8	ADN01129	Adn01129 Human cel
7	42	60.0	262	3	AAG43457	Aag43457 Arabidops
8	42	60.0	262	3	AAG23843	Aag23843 Arabidops
9	42	60.0	312	6	ABU23840	Abu23840 Protein e

10	42	60.0	344	7	ADC87149	Adc87149 Human GPC
11	42	60.0	351	3	AAG23842	Aag23842 Arabidops
12	42	60.0	351	3	AAG43456	Aag43456 Arabidops
13	42	60.0	381	3	AAG43455	Aag43455 Arabidops
14	42	60.0	381	3	AAG23841	Aag23841 Arabidops
15	41	58.6	90	6	ABU61687	Abu61687 Program d
16	41	58.6	90	7	ADB66818	Adb66818 Potassium
17	41	58.6	348	1	AAP94682	Aap94682 Human man
18	41	58.6	648	8	ADQ59397	Adq59397 Human can
19	41	58.6	856	7	ABW01409	Abw01409 Human HEA
20	41	58.6	962	3	AAY49944	Aay49944 Human pot
21	41	58.6	988	4	AAB31714	Aab31714 A human a
22	41	58.6	988	4	AAE01043	Aae01043 Human eth
23	41	58.6	988	5	ABG70999	Abg70999 Human tra
24	41	58.6	988	5	ABB76165	Abb76165 Human pot
25	41	58.6	988	5	AAO14214	Aao14214 Human tra
26	41	58.6	988	7	ADC77655	Adc77655 Human 334
27	41	58.6	988	7	ABW01364	Abw01364 Human pot
28	41	58.6	988	8	ADQ59400	Adq59400 Human can
29	41	58.6	989	3	AAY49945	Aay49945 Human pot
30	41	58.6	989	7	ADC77667	Adc77667 Human 332
31	41	58.6	989	8	ADR44914	Adr44914 Polypepti
32	41	58.6	1174	4	ABB65304	Abb65304 Drosophil
33	40	57.1	59	8	ABO56119	Abo56119 Human gen
34	40	57.1	84	4	AAU63085	Aau63085 Propionib
35	40	57.1	84	6	ABM59604	Abm59604 Propionib
36	40	57.1	699	8	ABO84631	Abo84631 Mouse can
37	40	57.1	1178	8	ADP74025	Adp74025 Murine CD
38	39	55.7	85	4	AAU21335	Aau21335 Human nov
39	39	55.7	103	4	AAU59241	Aau59241 Propionib
40	39	55.7	103	6	ABM55760	Abm55760 Propionib
41	39	55.7	166	7	ABO75576	Abo75576 Pseudomon
42	39	55.7	194	3	AAG61086	Aag61086 Arabidops
43	39	55.7	194	3	AAG56099	Aag56099 Arabidops
44	39	55.7	198	3	AAG56098	Aag56098 Arabidops
45	39	55.7	198	3	AAG61085	Aag61085 Arabidops

#### ALIGNMENTS

##### RESULT 1

AAB95737

ID AAB95737 standard; protein; 121 AA.

XX

AC AAB95737;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:18627.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 18627; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 121 AA;  
  
Query Match 62.9%; Score 44; DB 4; Length 121;  
Best Local Similarity 66.7%; Pred. No. 28;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 MCPTDMPASLCM 12  
||| :|||||  
Db 76 MCPVPVPASLSM 87

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2005, 10:55:53 ; Search time 41 Seconds  
(without alignments)  
21.849 Million cell updates/sec

Title: US-10-618-336A-6

Perfect score: 70

Sequence: 1 MCPTDMPASLCM 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

%

Result	Query					Description
No.	Score	Match	Length	DB	ID	
<hr/>						
1	41	58.6	90	4	US-09-358-383C-31	Sequence 31, Appl
2	41	58.6	960	4	US-09-694-777A-21	Sequence 21, Appl
3	41	58.6	962	4	US-09-614-480-9	Sequence 9, Appli
4	41	58.6	962	4	US-09-694-777A-3	Sequence 3, Appli
5	41	58.6	962	4	US-09-694-777A-24	Sequence 24, Appl
6	41	58.6	962	4	US-10-422-075-9	Sequence 9, Appli
7	41	58.6	987	4	US-09-694-777A-22	Sequence 22, Appl
8	41	58.6	988	4	US-09-614-480-2	Sequence 2, Appli
9	41	58.6	988	4	US-10-162-012-5	Sequence 5, Appli
10	41	58.6	988	4	US-10-162-012-12	Sequence 12, Appl
11	41	58.6	988	4	US-10-422-075-2	Sequence 2, Appli

12	41	58.6	989	4	US-09-694-777A-4	Sequence 4, Appli
13	41	58.6	989	4	US-09-694-777A-23	Sequence 23, Appl
14	40	57.1	96	4	US-09-403-343B-24	Sequence 24, Appl
15	39	55.7	166	4	US-09-252-991A-24322	Sequence 24322, A
16	38	54.3	103	4	US-09-252-991A-28978	Sequence 28978, A
17	38	54.3	2500	2	US-08-801-263A-2	Sequence 2, Appli
18	38	54.3	2500	3	US-09-102-248-2	Sequence 2, Appli
19	38	54.3	2500	4	US-09-367-764-2	Sequence 2, Appli
20	38	54.3	2517	2	US-08-801-263A-5	Sequence 5, Appli
21	38	54.3	2517	3	US-09-102-248-5	Sequence 5, Appli
22	38	54.3	2517	4	US-09-367-764-5	Sequence 5, Appli
23	37	52.9	250	4	US-09-270-767-61069	Sequence 61069, A
24	37	52.9	433	4	US-09-270-767-45556	Sequence 45556, A
25	37	52.9	534	4	US-09-270-767-46023	Sequence 46023, A
26	37	52.9	3830	4	US-09-693-205A-4	Sequence 4, Appli
27	36	51.4	42	4	US-09-270-767-31926	Sequence 31926, A
28	36	51.4	42	4	US-09-270-767-47143	Sequence 47143, A
29	36	51.4	200	4	US-09-252-991A-18742	Sequence 18742, A
30	36	51.4	256	4	US-09-949-016-7576	Sequence 7576, Ap
31	36	51.4	256	4	US-09-949-016-7577	Sequence 7577, Ap
32	36	51.4	280	4	US-09-949-016-9249	Sequence 9249, Ap
33	36	51.4	280	4	US-09-949-016-9250	Sequence 9250, Ap
34	36	51.4	306	4	US-09-270-767-41758	Sequence 41758, A
35	36	51.4	330	3	US-09-188-930-125	Sequence 125, App
36	36	51.4	330	4	US-09-312-283C-125	Sequence 125, App
37	36	51.4	404	3	US-09-191-608-18	Sequence 18, Appl
38	36	51.4	431	4	US-09-270-767-57907	Sequence 57907, A
39	36	51.4	443	4	US-09-266-965-131	Sequence 131, App
40	36	51.4	447	3	US-09-191-608-19	Sequence 19, Appl
41	36	51.4	471	3	US-09-191-608-17	Sequence 17, Appl
42	36	51.4	497	3	US-09-191-608-20	Sequence 20, Appl
43	36	51.4	555	4	US-09-270-767-42597	Sequence 42597, A
44	36	51.4	617	3	US-09-188-930-303	Sequence 303, App
45	36	51.4	617	4	US-09-312-283C-303	Sequence 303, App

#### ALIGNMENTS

#### RESULT 1

US-09-358-383C-31

; Sequence 31, Application US/09358383C  
; Patent No. 6518398  
; GENERAL INFORMATION:  
; APPLICANT: Curtis, Rory A.J.  
; TITLE OF INVENTION: NOVEL POTASSIUM CHANNEL MOLECULES AND USES THEREFOR  
; FILE REFERENCE: MNI-055CP  
; CURRENT APPLICATION NUMBER: US/09/358,383C  
; CURRENT FILING DATE: 1999-07-21  
; PRIOR APPLICATION NUMBER: USSN 09/119,855  
; PRIOR FILING DATE: 1998-07-21  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 31  
; LENGTH: 90  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: domain  
US-09-358-383C-31

Query Match 58.6%; Score 41; DB 4; Length 90;  
Best Local Similarity 54.5%; Pred. No. 15;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CPTDMPASLCM 12  
|| || | :|:  
Db 57 CPKDMKADICV 67

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2005, 11:04:24 ; Search time 155 Seconds  
(without alignments)  
30.163 Million cell updates/sec

Title: US-10-618-336A-6

Perfect score: 70

Sequence: 1 MCPTDMPASLCM 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1745140 seqs, 389608008 residues

Total number of hits satisfying chosen parameters: 1745140

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*
   
 17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep:\*
   
 18: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
   
 19: /cgn2\_6/ptodata/2/pubpaa/US11A\_PUBCOMB.pep:\*
   
 20: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*
   
 21: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
   
 22: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

8

Result	Query					Description
No.	Score	Match	Length	DB	ID	
1	44	62.9	47	15	US-10-424-599-193456	Sequence 193456,
2	44	62.9	296	15	US-10-210-281-4	Sequence 4, Appli
3	44	62.9	307	15	US-10-210-281-6	Sequence 6, Appli
4	44	62.9	329	15	US-10-112-944-313	Sequence 313, App
5	44	62.9	351	15	US-10-333-572-3	Sequence 3, Appli
6	42	60.0	312	15	US-10-282-122A-51764	Sequence 51764, A
7	42	60.0	344	14	US-10-017-161-1954	Sequence 1954, Ap
8	42	60.0	344	15	US-10-292-798-1602	Sequence 1602, Ap
9	42	60.0	381	16	US-10-739-930-5642	Sequence 5642, Ap
10	41	58.6	90	14	US-10-185-867-31	Sequence 31, Appl
11	41	58.6	149	16	US-10-425-115-351203	Sequence 351203,
12	41	58.6	263	16	US-10-437-963-164408	Sequence 164408,
13	41	58.6	275	16	US-10-739-930-8285	Sequence 8285, Ap
14	41	58.6	648	16	US-10-322-696-33	Sequence 33, Appl
15	41	58.6	960	14	US-10-188-308-21	Sequence 21, Appl
16	41	58.6	960	14	US-10-188-296-21	Sequence 21, Appl
17	41	58.6	960	14	US-10-188-341-21	Sequence 21, Appl
18	41	58.6	960	14	US-10-188-297-21	Sequence 21, Appl
19	41	58.6	962	14	US-10-188-308-3	Sequence 3, Appli
20	41	58.6	962	14	US-10-188-308-24	Sequence 24, Appl
21	41	58.6	962	14	US-10-188-296-3	Sequence 3, Appli
22	41	58.6	962	14	US-10-188-296-24	Sequence 24, Appl
23	41	58.6	962	14	US-10-188-341-3	Sequence 3, Appli
24	41	58.6	962	14	US-10-188-341-24	Sequence 24, Appl
25	41	58.6	962	14	US-10-188-297-3	Sequence 3, Appli
26	41	58.6	962	14	US-10-188-297-24	Sequence 24, Appl
27	41	58.6	962	14	US-10-174-613-4	Sequence 4, Appli
28	41	58.6	962	14	US-10-174-613-5	Sequence 5, Appli
29	41	58.6	962	15	US-10-422-075-9	Sequence 9, Appli
30	41	58.6	987	14	US-10-188-308-22	Sequence 22, Appl
31	41	58.6	987	14	US-10-188-296-22	Sequence 22, Appl
32	41	58.6	987	14	US-10-188-341-22	Sequence 22, Appl
33	41	58.6	987	14	US-10-188-297-22	Sequence 22, Appl
34	41	58.6	988	10	US-09-875-321-5	Sequence 5, Appli
35	41	58.6	988	10	US-09-875-321-12	Sequence 12, Appl
36	41	58.6	988	14	US-10-128-323-2	Sequence 2, Appli
37	41	58.6	988	14	US-10-162-012-5	Sequence 5, Appli
38	41	58.6	988	14	US-10-162-012-12	Sequence 12, Appl
39	41	58.6	988	14	US-10-174-613-2	Sequence 2, Appli
40	41	58.6	988	14	US-10-174-613-3	Sequence 3, Appli

41	41	58.6	988	15	US-10-369-022-18	Sequence 18, Appl
42	41	58.6	988	15	US-10-422-075-2	Sequence 2, Appli
43	41	58.6	988	15	US-10-162-102-5	Sequence 5, Appli
44	41	58.6	988	15	US-10-162-102-12	Sequence 12, Appl
45	41	58.6	988	15	US-10-332-447-31	Sequence 31, Appl

### ALIGNMENTS

#### RESULT 1

US-10-424-599-193456

; Sequence 193456, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 193456

; LENGTH: 47

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_16715C.1.pep

US-10-424-599-193456

Query Match 62.9%; Score 44; DB 15; Length 47;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MCPTDMPASLCM 12  
 :| |:||: ||:

Db 9 LCQTNMPSKLCL 20

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2005, 10:55:03 ; Search time 38 Seconds  
 (without alignments)  
 30.384 Million cell updates/sec

Title: US-10-618-336A-6

Perfect score: 70

Sequence: 1 MCPTDMPASLCM 12

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*

1: pirl:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

8

Result No.	Score	Query				Description
		Match	Length	DB	ID	
1	47	67.1	3839	2	T49799	related to TOM1 pr
2	42	60.0	312	2	G97162	hypothetical prote
3	42	60.0	381	2	F84615	hypothetical prote
4	42	60.0	1206	2	E86445	hypothetical prote
5	41	58.6	763	2	A47563	glucose-6-phosphat
6	41	58.6	934	2	T42394	potassium channel
7	41	58.6	962	2	I53197	potassium channel
8	41	58.6	989	2	I48912	potassium channel
9	41	58.6	1174	2	A40853	potassium channel
10	40	57.1	471	2	AC2014	two-component sens
11	40	57.1	1178	2	S44142	VLA-2 protein homo
12	39	55.7	537	2	I51592	protein-tyrosine k
13	38.5	55.0	172	2	T34413	hypothetical prote
14	38	54.3	320	2	T09288	late embryonic abu
15	38	54.3	746	2	T43648	probable sexual cy
16	38	54.3	1461	2	T13157	mitotic checkpoint
17	37	52.9	182	2	D70859	hypothetical prote
18	37	52.9	388	2	T06035	hypothetical prote
19	37	52.9	398	1	TVFVVR	transforming prote
20	37	52.9	676	2	A45984	sperm-binding glyc
21	37	52.9	782	2	T22134	hypothetical prote
22	36.5	52.1	965	2	T12545	hypothetical prote
23	36	51.4	130	2	G86804	cytidine deaminase
24	36	51.4	249	2	H71131	hypothetical prote
25	36	51.4	384	2	AG3494	recF protein [impo
26	36	51.4	408	2	S33683	'site-specific DNA-
27	36	51.4	1611	2	T38236	hypothetical prote
28	36	51.4	1894	2	JC4980	plexin 1 precursor
29	36	51.4	2533	2	T28675	alpha-51D immobili
30	36	51.4	2533	2	T28674	alpha-51D-immobili

31	35.5	50.7	561	2	T11378	NADH2 dehydrogenas
32	35	50.0	131	2	AG3198	hypothetical prote
33	35	50.0	138	2	T48293	hypothetical prote
34	35	50.0	220	2	T30688	hypothetical prote
35	35	50.0	237	2	T24407	hypothetical prote
36	35	50.0	246	2	S74408	lipopeptide antibi
37	35	50.0	321	2	E97741	D-alanine-D-alanin
38	35	50.0	360	2	T35783	probable secreted
39	35	50.0	365	2	T43286	cet-1 protein - Ca
40	35	50.0	377	2	AD2922	altronate oxidored
41	35	50.0	377	2	E97696	altronate oxidored
42	35	50.0	379	2	T29518	hypothetical prote
43	35	50.0	432	2	T23762	hypothetical prote
44	35	50.0	461	2	T41369	hypothetical prote
45	35	50.0	466	2	F69331	conserved hypothet

#### ALIGNMENTS

#### RESULT 1

T49799

related to TOM1 protein [imported] - Neurospora crassa

N;Alternate names: protein B11B22.10

C;Species: Neurospora crassa

C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004

C;Accession: T49799

R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.;

Nyakatura, G.; Mewes, H.W.; Mannhaupt, G.

submitted to the Protein Sequence Database, May 2000

A;Reference number: Z25022

A;Accession: T49799

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-3839 <SCH>

A;Cross-references: UNIPROT:Q9P4Z1; EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.10

A;Experimental source: BAC clone B11B22; strain OR74A

C;Genetics:

A;Gene: NCSP:B11B22.10

A;Map position: 6

A;Introns: 16/3; 2607/1; 2623/1; 2658/1; 2845/1; 2987/2; 3204/3; 3694/1; 3809/1

Query Match 67.1%; Score 47; DB 2; Length 3839;  
 Best Local Similarity 70.0%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCPTDMPASL 10  
 ||||:|||||:  
 Db 381 MCPSDMPATM 390

GenCore version 5.1.6  
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2005, 10:54:23 ; Search time 174 Seconds  
(without alignments)  
35.316 Million cell updates/sec

Title: US-10-618-336A-6  
Perfect score: 70  
Sequence: 1 MCPTDMPASLCM 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query						Description
No.	Score	Match	Length	DB	ID		
1	47	67.1	4065	1	TOM1_NEUCR		Q9p4z1 neurospora
2	44	62.9	121	2	Q9H7W8		Q9h7w8 homo sapien
3	44	62.9	365	2	Q8IV13		Q8iv13 homo sapien
4	42	60.0	312	1	MRAW_CLOAB		Q97h81 clostridium
5	42	60.0	381	2	Q8LA81		Q8la81 arabidopsis
6	42	60.0	381	2	Q9ZQ45		Q9zq45 arabidopsis
7	42	60.0	1206	2	Q9FVQ7		Q9fvq7 arabidopsis
8	41	58.6	27	2	Q81172		Q81172 hepatitis b
9	41	58.6	204	2	Q986U1		Q986u1 rhizobium l
10	41	58.6	278	2	Q6Z2H6		Q6z2h6 oryza sativ
11	41	58.6	297	2	Q7QNM1		Q7qnm1 anopheles g
12	41	58.6	543	2	Q9W645		Q9w645 gallus gall
13	41	58.6	568	2	Q847W2		Q847w2 vibrio para
14	41	58.6	599	2	Q87PX3		Q87px3 vibrio para
15	41	58.6	763	1	G6PE_RABIT		P56201 oryctolagus
16	41	58.6	956	2	O44164		O44164 caenorhabdi
17	41	58.6	956	2	Q9XYX7		Q9xyx7 caenorhabdi
18	41	58.6	962	1	KCH1_RAT		Q63472 rattus norv
19	41	58.6	987	1	KCH1_BOVIN		O18965 bos taurus
20	41	58.6	988	1	KCH5_HUMAN		Q8ncm2 homo sapien
21	41	58.6	988	1	KCH5_RAT		Q9epi9 rattus norv
22	41	58.6	988	2	Q6NZH0		Q6nzh0 mus musculu

23	41	58.6	988	2	Q8C035	Q8c035	mus musculu
24	41	58.6	989	1	KCH1_HUMAN	O95259	homo sapien
25	41	58.6	989	1	KCH1_MOUSE	Q60603	mus musculu
26	41	58.6	1007	2	Q7QCM9	Q7qcm9	anopheles g
27	41	58.6	1011	2	Q7YW98	Q7yw98	manduca sex
28	41	58.6	1174	1	CIKE_DROME	Q02280	drosophila
29	41	58.6	1174	2	Q9VXZ6	Q9vxz6	drosophila
30	41	58.6	1285	2	Q8JUV1	Q8juv1	avian reovi
31	40	57.1	323	2	Q8CB84	Q8cb84	mus musculu
32	40	57.1	471	2	Q8YWE8	Q8ywe8	anabaena sp
33	40	57.1	508	1	VL1 HPV60	P50821	human papil
34	40	57.1	1178	1	ITA2_MOUSE	Q62469	mus musculu
35	40	57.1	1178	2	Q6P1C7	Q6p1c7	mus musculu
36	40	57.1	2116	2	Q6D9Y1	Q6d9y1	erwinia car
37	40	57.1	2168	2	Q6LGA3	Q6lga3	photobacter
38	39	55.7	196	2	Q6Y2V5	Q6y2v5	desulfuromo
39	39	55.7	198	2	Q8LF87	Q8lf87	arabidopsis
40	39	55.7	222	2	Q9LT60	Q9lt60	arabidopsis
41	39	55.7	264	2	Q66IP2	Q66ip2	xenopus lae
42	39	55.7	300	2	Q9YUQ5	Q9yuq5	turkey aden
43	39	55.7	318	2	Q67Q57	Q67q57	symbiobacte
44	39	55.7	328	2	Q8LJD0	Q81jd0	oryza sativ
45	39	55.7	346	2	Q652V2	Q652v2	oryza sativ

#### ALIGNMENTS

#### RESULT 1

TOM1\_NEUCR  
 ID TOM1\_NEUCR STANDARD; PRT; 4065 AA.  
 AC Q9P4Z1;  
 DT 25-OCT-2004 (Rel. 45, Created)  
 DT 25-OCT-2004 (Rel. 45, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE E3 ubiquitin protein ligase TOM1-like protein (EC 6.3.2.-).  
 GN ORFNames=B11B22.010, NCU08501.1;  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A / FGSC 987;  
 RX MEDLINE=22542210; PubMed=12655011; DOI=10.1093/nar/gkg293;  
 RA Mannhaupt G., Montrone C., Haase D., Mewes H.-W., Aign V.,  
 RA Hoheisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,  
 RA Schulte U.;  
 RT "What's in the genome of a filamentous fungus? Analysis of the  
 RT Neurospora genome sequence.";  
 RL Nucleic Acids Res. 31:1944-1954(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A / FGSC 987;  
 RX PubMed=12712197; DOI=10.1038/nature01554;  
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,  
 RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,

RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,  
RA Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A.,  
RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,  
RA Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,  
RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,  
RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysselis M.,  
RA Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S.,  
RA Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S., Cogoni C.,  
RA Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,  
RA DeSouza C.P., Glass L., Orbach M.J., Berglund J.A., Voelker R.,  
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebole D.J., Freitag M.,  
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;  
RT "The genome sequence of the filamentous fungus Neurospora crassa.";  
RL Nature 422:859-868 (2003).

CC -!- FUNCTION: Probable ubiquitin ligase protein, which may be involved  
CC in mRNA export. E3 ubiquitin ligase protein mediate ubiquitination  
CC and subsequent proteasomal degradation of target proteins.  
CC Participates in mRNA export from the nucleus by regulating the  
CC transport of hnRNP proteins (By similarity).  
CC -!- PATHWAY: Ubiquitin conjugation; third step.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: Belongs to the TOM1/PTR1 family.  
CC -!- SIMILARITY: Contains 1 HECT-type E3 ubiquitin-protein ligase  
CC domain.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to license@isb-sib.ch).  
CC -----

DR EMBL; AL356834; CAB92704.2; -.  
DR EMBL; AABX01000134; EAA34194.1; -.  
DR PIR; T49799; T49799;  
DR HSSP; Q05086; 1C4Z.  
DR InterPro; IPR010309; DUF908.  
DR InterPro; IPR010314; DUF913.  
DR InterPro; IPR000569; HECT\_domain.  
DR Pfam; PF06012; DUF908; 1.  
DR Pfam; PF06025; DUF913; 1.  
DR Pfam; PF00632; HECT; 1.  
DR SMART; SM00119; HECTc; 1.  
DR PROSITE; PS50237; HECT; 1.

KW Coiled coil; Ligase; mRNA transport; Nuclear protein; Transport;  
KW Ubl conjugation pathway.

FT DOMAIN 3729 4065 HECT.  
FT DOMAIN 1263 1287 Coiled coil (Potential).  
FT DOMAIN 2491 2568 Coiled coil (Potential).  
FT DOMAIN 2839 2924 Coiled coil (Potential).  
FT DOMAIN 3320 3369 Coiled coil (Potential).  
FT BINDING 4032 4032 Ubiquitin (By similarity).  
SQ SEQUENCE 4065 AA; 452568 MW; F74683CEC36F9350 CRC64;

Query Match 67.18; Score 47; DB 1; Length 4065;

Best Local Similarity 70.0%; Pred. No. 65;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCPTDMPASL 10  
|||:||||:|:  
Db 381 MCPSDMPATM 390